

# Reshaped tensor nuclear norms for higher order tensor completion

Kishan Wimalawarne<sup>1</sup> 💿 • Hiroshi Mamitsuka<sup>2,3</sup>

Received: 11 June 2020 / Revised: 11 June 2020 / Accepted: 21 October 2020 / Published online: 3 January 2021 © The Author(s), under exclusive licence to Springer Science+Business Media LLC, part of Springer Nature 2021

## Abstract

We investigate optimal conditions for inducing low-rankness of higher order tensors by using convex tensor norms with reshaped tensors. We propose the reshaped tensor nuclear norm as a generalized approach to reshape tensors to be regularized by using the tensor nuclear norm. Furthermore, we propose the reshaped latent tensor nuclear norm to combine multiple reshaped tensors using the tensor nuclear norm. We analyze the generalization bounds for tensor completion models regularized by the proposed norms and show that the novel reshaping norms lead to lower Rademacher complexities. Through simulation and real-data experiments, we show that our proposed methods are favorably compared to existing tensor norms consolidating our theoretical claims.

Keywords Tensor nuclear norm · Reshaping · CP rank · Generalization bounds

# 1 Introduction

Tensor formatted data is becoming abundant in machine learning applications. Among the many tensor related machine learning problems, tensor completion has gained an increased popularity in recent years. Tensor completion performs imputation of unknown elements of a partially observed tensor by exploiting its low-rank structure. Some of the popular real-world applications of tensor completion are found in recommendation systems (Karatzoglou et al. 2010; Zheng et al. 2010), computer vision (Liu et al. 2009), and multi-relational link prediction (Rai et al. 2015). Though there exist many methods to perform tensor completion (Song

Editor: Pradeep Ravikumar.

Kishan Wimalawarne kishanwn@gmail.com

Hiroshi Mamitsuka mami@kuicr.kyoto-u.ac.jp

<sup>2</sup> Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Japan

<sup>&</sup>lt;sup>1</sup> Department of Mathematical Informatics, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-8656, Japan

<sup>&</sup>lt;sup>3</sup> Department of Computer Science, Aalto University, Espoo, Finland

et al. 2017), global optimal solutions are obtained mainly by convex low-rank tensor norms, making them an active area of research.

Over the years, many researchers have proposed different low-rank inducing norms to minimize the rank of tensors, however, none of these norms are universally better compared to others. The main challenge in designing norms for tensors is that they have multiple dimensions and different definitions of ranks (Tucker rank, CP rank, TT-rank), making it difficult for a single norm to induce low-rankness with respect to all the properties of tensors. Most tensor norms have been designed with a focus to a specific rank; overlapped trace norm (Tomioka and Suzuki 2013) and latent trace norms (Wimalawarne et al. 2014) to constrain the multilinear ranks, tensor nuclear norm (Yuan and Zhang 2016; Yang et al. 2015; Lim and Comon 2014) to constrain the CP rank, and the Schatten TT rank (Imaizumi et al. 2017) to constrain the TT-rank. However, targeting a specific rank to constrain may not always be practical, since we may not know the most suitable rank for a tensor in advance.

Most tensor norms reshape tensors by rearranging its elements as matrices to induce lowrankness with respect to a mode or a set of modes. However, this reshaping method is specific to obtaining relevant ranks that a norm constrains. An alternative view was presented by Mu et al. (2014) with the square norm, where the tensor is reshaped as a balanced matrix without considering the structure of its ranks. The square norm has shown to have better sample complexities for higher order tensors (tensor with more than three modes) than some of the existing norms such as the overlapped trace norm (Yuan and Zhang 2016). However, this norm only considers the special case of reshaping a tensor as a matrix such that its dimensions are close to each other. Other possibilities of how reshaping tensors beyond matrices affect the inducement of low-rankness have not been investigated.

In this paper, we propose generalized reshaping strategies to reshape tensors and develop low-rank inducing tensor norms. We demonstrate that reshaping a higher order tensor as another tensor and applying the tensor nuclear norm leads to better inducement of low-rankness compared to applying existing low-rank norms on the original tensor or its matrix unfoldings. Furthermore, we propose the latent reshaped tensor nuclear norm that combines multiple reshaped tensors to obtain a better performance among a set of reshaped tensors. Using the generalization bounds, we show that the proposed norms are able to give lower Rademacher complexities compared to existing norms. Using simulations and real-world data experiments we justify our theoretical analysis and show that our proposed methods are able to give better performances for tensor completion compared other convex norms.

Throughout this paper we use the following notations. We represent a *K*-mode (*K*-way) tensor as  $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ . The mode-*k* unfolding (Kolda and Bader 2009) of a tensor  $\mathcal{T}$  is given by  $T_{(k)} \in \mathbb{R}^{n_k \times \prod_{j \neq k} n_j}$ , which is obtained by concatenating all slices along the mode-*k*. We indicate the tensor product (Hackbusch 2012) between vectors  $u_i \in \mathbb{R}^{n_i}$ ,  $i = 1, \ldots, K$  using the notation  $\otimes$  as  $(u_1 \otimes \cdots \otimes u_K)_{i_1,\ldots,i_K} = \prod_{l=1}^K u_{l,i_l}$ . The *k*-mode product of a tensor  $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_k \cdots \times n_K}$  and a vector  $v \in \mathbb{R}^{n_k}$  is defined as  $\mathcal{T} \times_k v = \sum_{i_k=1}^{n_k} \mathcal{I}_{i_1,i_2,\ldots,i_k, \dots,i_K} v_{i_k}$ . The largest singular value of  $\mathcal{T}$  is given by  $\gamma_1(\mathcal{T})$ . The rank of a matrix  $A \in \mathbb{R}^{n \times m}$  is given by Rank(A).

# 2 Review of low-rank tensor norms

Designing of convex low-rank inducing norms for tensors is a challenging task. Over the years, several tensor norms have been proposed with each norm having certain advantages over the others. The main challenge with defining tensor norms is the multidimensionality of tensors and the existence of different ranks (e.g. CP rank, multilinear (Tucker) rank). A common criterion for designing low-rank tensor norms is to induce low-rankness by minimizing a particular rank. A commonly used rank is the multilinear rank, which represents the rank with respect to each mode of a tensor. Given a tensor  $W \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ , we obtain the rank of each unfolding  $r_k = \operatorname{Rank}(W_{(k)})$ ,  $k = 1, \ldots, K$ , and define the multilinear rank as  $(r_1, \ldots, r_K)$ . To minimize the multilinear rank the overlapped trace norm has been defined (Liu et al. 2009; Tomioka and Suzuki 2013), which for a tensor  $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ , is

$$\|\mathcal{T}\|_{\text{overlap}} = \sum_{k=1}^{K} \|T_{(k)}\|_{\text{tr}},$$

where  $\|\cdot\|_{tr}$  is the matrix nuclear norm (a.k.a. trace norm) (Fazel et al. 2001), which is the sum of the non-zero singular values of a matrix. A limitation with this norm is that for tensors with high variations in the multilinear rank this norm stays at poor performances (Tomioka and Suzuki 2013; Wimalawarne et al. 2014).

The latent trace norm (Tomioka and Suzuki 2013) has been proposed to overcome limitations of the overlapped trace norm, which allows freedom to learn ranks with respect to each mode unfolding by considering a latent decomposition of the tensor. More specifically, the latent tensor norm learns latent tensors  $T^{(k)}$ , k = 1, ..., K as

$$\|\mathcal{T}\|_{\text{latent}} = \inf_{\mathcal{T}^{(1)} + \ldots + \mathcal{T}^{(k)} = \mathcal{T}_{k=1}^{k}} \|\mathcal{T}_{(k)}^{(k)}\|_{\text{tr}}.$$

This norm was shown to be more robust for tensors with high variations in the multilinear rank compared to the overlapped trace norm (Tomioka and Suzuki 2013). The latent trace norm has been further extended to develop the scaled latent trace norm (Wimalawarne et al. 2014) by considering the relative rank of each latent tensor by scaling using the inverse squared mode dimension.

Another popular rank for tensors is the CANDECOMP/PARAFAC (CP) rank (Carroll and Chang 1970; Harshman 1970; Hitchcock 1927; Kolda and Bader 2009), which can be considered as the higher order extension of the matrix rank. Recently, minimization of the CP rank has gained attention of many researchers, who have shown that it leads to a better sample complexity than multilinear rank based norms (Yuan and Zhang 2016). The tensor nuclear norm (Yuan and Zhang 2016; Yang et al. 2015; Lim and Comon 2014) has been defined as an approximation to minimize the CP rank of a tensor. For a tensor  $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$  with rank *R*, Rank( $\mathcal{T}$ ) = *R*, the tensor nuclear norm is defined as

$$\|\mathcal{T}\|_{*} = \inf \left\{ \sum_{j=1}^{R} \gamma_{j} | \mathcal{T} = \sum_{j=1}^{R} \gamma_{j} u_{1j} \otimes u_{2j} \cdots \otimes u_{Kj}, \\ \|u_{kj}\|_{2}^{2} = 1, \gamma_{j} \ge \gamma_{j+1} > 0 \right\},$$
(1)

where  $u_{ki} \in \mathbb{R}^{n_k}$  for  $k = 1, \dots, K$  and  $j = 1, \dots, R$ .

The latest addition to convex low-rank tensor norms is the Schatten TT norm (Imaizumi et al. 2017), which minimizes the tensor train rank (Oseledets 2011) of tensors. The Schatten TT norm is defined as

$$\|\mathcal{T}\|_{s,T} = \frac{1}{K-1} \sum_{k=1}^{K-1} \|\mathcal{Q}_k(\mathcal{T})\|_{\mathrm{tr}},$$

where  $Q_k : \mathcal{T} \to \mathbb{R}^{n_{\geq k} \times n_{k <}}$  is an operator that reshapes the tensor  $\mathcal{T}$  to a matrix by combining the first k modes as rows and the rest of the K - k modes as columns. This norm has been shown to be suitable for high-order tensors.

It has also been shown that low-rank tensor norms can be designed without restricting to a specific rank. The square norm (Mu et al. 2014) reshapes a tensor as a matrix and apply the matrix nuclear norm as

$$\left\|\mathcal{T}\right\|_{\Box} = \left\|\operatorname{reshape}\left(\mathcal{T}_{(1)}, \prod_{i=1}^{j}, \prod_{i=j+1}^{K}\right)\right\|_{\mathrm{tr}},$$

where the function reshape() reshapes T to a matrix with approximately equal dimensions for some j > 1. This norm has shown to have a better sample complexity for tensor completion compared to the overlapped trace norm.

We point out that all the existing tensor norms except the tensor nuclear norm reshape tensors as matrices to induce the low-rankness with respect to two sets of mode arrangements. As a result these norms are focused on constraining the multilinear rank of a tensor. However, tensor nuclear norm has shown to lead to a better sample complexity compared to multilinear rank based tensors norms for tensor completion (Yuan and Zhang 2016). Hence, lack of tensor nuclear norm regularization for reshaped tensors among existing norms may results in sub-optimal solutions.

### 3 Proposed method: tensor reshaping and tensor nuclear norm

In this paper, we investigate on extending the tensor nuclear norm for higher order tensors. We explore methods to combine tensor reshaping with the tensor nuclear norm.

#### 3.1 Generalized tensor reshaping

First, we introduce the following notation to compute the product of tensor dimensions. For a given vector  $(n_1, \ldots, n_p)$ , we represent its element-wise product by  $prod(n_1, \ldots, n_p) = n_1 n_2 \cdots n_p$ . Next, we define generalized reshaping for tensors.

**Definition 1** (Tensor Reshaping) Let us consider a tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$  and its mode dimensions as  $D = \{n_1, n_2, \dots, n_K\}$ . Given M sets  $D_i \subset D$ ,  $i = 1, \dots, M$ , that are disjoint,  $D_i \cap D_i = \emptyset$  for  $i \neq j$ , the reshaping operator is defined as

$$\Pi_{(D_1,\ldots,D_M)}: \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K} \to \mathbb{R}^{\mathsf{prod}(D_1) \times \cdots \times \mathsf{prod}(D_M)}$$

and the inverse operator is represented by  $\Pi_{(D_1,...,D_M)}^{\top}$ . Further, we present the reshaping of  $\mathcal{X}$  by the set  $(D_1, \ldots, D_M)$  as  $\mathcal{X}_{(D_1,...,D_M)}$ .

We point out that when  $|D_1| = \cdots = |D_M| = 1$ , there is no reshaping of the tensor,  $\mathcal{X}_{(D_1,\dots,D_M)} = \mathcal{X}$ . Unfolding of a tensor along the mode *k* (Kolda and Bader 2009) is equivalent to defining two sets with  $D_1 = n_k$  and  $D_2 = (n_1, \dots, n_{k-1}, n_{k+1}, \dots, n_K)$ . Further, we can obtain reshaping of a tensor as a matrix for the square norm (Mu et al. 2014) by specifying two sets  $D_1$  and  $D_2$  with  $prod(D_1) \approx prod(D_2)$ .

#### 3.2 Reshaped tensor nuclear norm

We propose a class of tensor norms by combining generalized tensor reshaping and the tensor nuclear norm. We name the proposed norms *Reshaped Tensor Nuclear Norms*. In order to define the proposed norms, we consider a *K*-mode tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$  and a set  $D_i$ ,  $i = 1, \dots, M$ , adhering to Definition 1. We define the reshaped tensor nuclear norm as

$$\|\mathcal{X}_{(D_1,...,D_M)}\|_{*}$$

where  $\|\cdot\|_*$  is the tensor nuclear norm as defined in (1). It is understood that this norm is a convex norm, since the tensor nuclear norm (1) is convex.

#### 3.3 Reshaped latent nuclear norm

A practical limitation in applying reshaping using the proposed tensor norm is the difficulty to select the most suitable reshaping set out of all possible reshaping combinations. This is critical since we would not know the ranks of the tensor prior to training a learning model. To overcome this difficulty we propose the *Reshaped Latent Tensor Nuclear Norm* by extending the latent trace norm (Tomioka and Suzuki 2013) for reshaping tensors.

Let us consider a collection of G reshaping sets  $D_{\rm L} = (D^{(1)}, \dots, D^{(\bar{G})})$  where each  $D^{(s)} = (D_1^{(s)}, \dots, D_{m_s}^{(s)})$ ,  $s = 1, \dots, G$  consists a reshaping set for a  $m_s$ -mode reshaped tensor. Further, we consider the  $\mathcal{W}$  as a summation of G latent tensors  $\mathcal{W}^{(g)}$ ,  $g = 1, \dots, G$  as  $\mathcal{W} = \sum_{k=1}^{G} \mathcal{W}^{(k)}$ . We define the reshaped latent tensor nuclear norm as

$$\|\mathcal{W}\|_{r\_latent(D_L)} = \inf_{\mathcal{W}^{(1)} + \dots + \mathcal{W}^{(G)} = \mathcal{W}} \sum_{k=1}^{G} \|\mathcal{W}_{(D_1^{(k)}, \dots, D_{m_k}^{(k)})}^{(k)}\|_*.$$
 (2)

We point out that the above norm differs from the latent trace norm (Tomioka and Suzuki 2013) since it considers reshaping sets defined by the user where the latent trace norm considers all the mode-wise tensor unfolding. Furthermore, the above norm uses the tensor nuclear norm while the latent trace norm is build using the matrix nuclear norm.

#### 3.4 Completion models

Now, we propose tensor completion models for the proposed norms. Let us consider a partially observed tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ . Given that  $\mathcal{X}$  has *m* observed elements, we define the mapping of the observed elements from  $\mathcal{X}$  by  $\Omega : \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K} \to \mathbb{R}^m$ . Given a reshaping set  $(D_1, \ldots, D_M)$ , the completion model that is regularized by the reshaped norm is given as

$$\begin{split} \min_{\mathcal{W}} & \frac{1}{2} \| \mathcal{Q}(\mathcal{X}) - \mathcal{Q}(\mathcal{W}) \|_{\mathrm{F}}^{2} \\ \text{s.t.} & \| \mathcal{W}_{(D_{1}, \dots, D_{M})} \|_{*} \leq \lambda, \end{split}$$
(3)

where  $\lambda$  is a regularization parameter. For a selected set of reshaping sets  $D_{\rm L} = (D^{(1)}, \dots, D^{(G)})$ , a completion model regularized by the reshaped latent tensor nuclear norm is given as

$$\min_{\mathcal{W}^{(1)}+\dots+\mathcal{W}^{(G)}=\mathcal{W}} \frac{1}{2} \| \Omega(\mathcal{X}) - \Omega(\mathcal{W}^{(1)}+\dots+\mathcal{W}^{(G)}) \|_{\mathrm{F}}^{2}$$
s.t.  $\| \mathcal{W} \|_{\mathrm{r\_latent}(D_{\mathrm{L}})} \leq \lambda,$ 
(4)

where  $\lambda$  is a regularization parameter.

## 4 Theory

We investigate theoretical properties of our proposed methods to identify the optimal conditions for reshaping of tensors. For our analysis, we use generalization bounds based on the transductive Rademacher complexity analysis (El-Yaniv and Pechyony 2007; Shamir and Shalev-Shwartz 2014).

We consider the learning problem in (3) and we denote the indexes of the observed elements of  $\mathcal{X}$  by S, where each index  $(i_1, \dots, i_K)$  of observed elements of  $\mathcal{X}$  is assigned as an element  $\alpha_j \in S$  for some  $1 \le j \le |S|$ . We consider observed elements as the training set denoted by  $S_{\text{Train}}$  and the rest belonging to the test set denoted by  $S_{\text{Test}}$ . For the convenience of deriving the Rademacher complexity, we consider the special case of  $|S_{\text{Train}}| = |S_{\text{Test}}| = |S|/2$  as in (Shamir and Shalev-Shwartz 2014).

Given a reshaping set  $(D_1, ..., D_M)$ , we consider the hypothesis class  $W = \{W | || W_{(D_1,...,D_M)} ||_* \le t\}$  for a given t. Given a loss function  $l(\cdot, \cdot)$  and a set S, we define the empirical loss as

$$L_{\mathbf{S}}(l \circ \mathcal{W}) := \frac{1}{|\mathbf{S}|} \left[ \sum_{(i_1, \dots, i_K) \in \mathbf{S}} l(\mathcal{X}_{i_1, \dots, i_K}, \mathcal{W}_{i_1, \dots, i_K}) \right].$$

Given that  $\max_{i_1,\ldots,i_K W \in W} |l(\mathcal{X}_{i_1,\ldots,i_K}, \mathcal{W}_{i_1,\ldots,i_K})| \le b_l$ , it is straight forward to extend generalizing bounds for matrices from (Shamir and Shalev-Shwartz 2014) to tensors, which holds with probability  $1 - \delta$  as

$$\begin{split} L_{\mathrm{S}_{\mathrm{Test}}}(l \circ \mathcal{W}) - L_{\mathrm{S}_{\mathrm{Tmin}}}(l \circ \mathcal{W}) &\leq 4R_{\mathrm{S}}(l \circ \mathcal{W}) \\ &+ b_l \bigg( \frac{11 + 4\sqrt{\log \frac{1}{\delta}}}{\sqrt{|\mathrm{S}_{\mathrm{Tmin}}|}} \bigg), \end{split}$$

where  $R_{\rm S}(l_{\rm OW})$  is the transductive Rademacher complexity theory (El-Yaniv and Pechyony 2007; Shamir and Shalev-Shwartz 2014) defined as

$$R_{\rm S}(l\circ\mathcal{W}) = \frac{1}{|{\rm S}|} \mathbb{E}_{\sigma} \left[ \sup_{\mathcal{W}\in\mathcal{W}} \sum_{j=1}^{|{\rm S}|} \sigma_j l(\mathcal{X}_{\alpha j}, \mathcal{W}_{\alpha j}) \right],\tag{5}$$

where  $\sigma_j \in \{-1, 1\}, j = 1, ..., |S|$  with probability of 0.5 are Rademacher variables.

The following theorem gives the Rademacher complexity for completion models regularized by a reshaped tensor nuclear norm. **Theorem 1** Consider a K-mode tensor  $W \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ . Let us consider any M reshaping sets  $(D_1, \dots, D_M)$  with a hypothesis class of  $W = \{W | || W_{(D_1, \dots, D_M)} ||_* \le t\}$ . Suppose that for all  $(i_1, \dots, i_K)$ ,  $l(\mathcal{X}_{i_1, \dots, i_K}, \cdot)$  is A-Lipschitz continuous. Then,

(a) given that W has a multilinear rank of  $(r_1, \ldots, r_K)$ , we obtain

$$R_{S}(l \circ \mathcal{W}) \leq \frac{cA}{|S|} \left( \frac{\prod_{k=1}^{K} r_{k}}{\max_{j=1,...,M} \prod_{i \in D_{j}} r_{i}} \right) \gamma_{1}(\mathcal{W}_{(D_{1},...,D_{M})})$$
$$\log(4M) \sum_{j=1}^{M} \sqrt{\prod_{p \in D_{j}} n_{p}},$$

(b) given that W has a CP rank of  $r_{cp}$ , we obtain

$$R_{\mathbf{S}}(l \circ \mathcal{W}) \leq \frac{c\Lambda}{|\mathbf{S}|} r_{cp} \gamma_1(\mathcal{W}_{(D_1,\dots,D_M)}) \log(4M) \sum_{j=1}^M \sqrt{\prod_{p \in D_j} n_p},$$

where c is a constant.

Using the Theorem 1, we can obtain the Rademacher complexities for tensor nuclear norm by considering  $|D_1| = |D_2| = \cdots = |D_K| = 1$  and the square norm by two reshaping sets of  $|D_1|$  and  $|D_2|$  such that  $\prod_{p \in D_1} n_p \approx \prod_{q \in D_2} n_q$ . We summarize Rademacher complexities of convex low-rank tensor norms in Table 1 for a tensor with equal mode dimensions  $(n_1 = n_2 = \ldots = n_K = n)$ .

From Table 1 and Theorem 1, we see that norms constructed using the tensor nuclear norm lead to better bounds compared to the overlapped trace norm, latent trace norm, and the scaled latent trace norm. Further, we see that the mode based components of the Rademacher complexity would have the smallest value with the tensor nuclear norm  $(\log(4K)\sqrt{Kn})$ . It is also clear that for any reshaping set, we find that  $\log(4K)\sqrt{Kn} \le \log(4M)\sum_{j=1}^{M}\sqrt{n^{|D_j|}}$ . This observation might lead us to conclude that the tensor nuclear norm is better than all the other norms. However, considering the multilinear rank such that  $1 < r_1 \le r_2 \le \cdots \le r_K$ , we can always find M < K reshaping sets  $D_1, D_2, \ldots, D_M$  such that  $\frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\ldots,K} r_j} \ge \frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\ldots,M} \prod_{i\in D_j} r_i}$ . In other words, we can reshape the tensor such that the Rademacher complexity for the reshaped tensor nuclear norm is bounded with a smaller rank based component compared to the tensor nuclear norm.

It is not known how reshaping a tensor changes the CP rank of the original tensor into the rank of the reshaped tensor except that  $\operatorname{Rank}(\mathcal{X}_{(D_1,\ldots,D_M)}) \leq r_{cp}$  ((Mu et al. 2014) and Lemma 4 in "Appendix"). However, Theorem 1 shows that reshaping results in a mode based component of  $\log(4M) \sum_{j=1}^{M} \sqrt{n^{|D_j|}}$  in the Rademacher complexity, which indicates that selecting a reshaping set that gives a lower mode based components can lead to a lower generalization bound compared the square norm or the tensor nuclear norm. Furthermore, it is clear that the reshaping a tensor and regularization using the tensor nuclear norm lead to a lower generalization bound compared to multilinear rank based norms such as the overlapped trace norm, latent trace norm, and scaled latent trace norms and tensor train rank based Schatten TT norm.

The next theorem provides the Rademacher complexity for completion models regularized by the reshaped latent tensor nuclear norm.

Rademacher complexity $R_{\rm S}(l \circ W)$
$\frac{\partial A}{\partial U} \sum_{i=1}^{K} \sqrt{r_i} B_{\mathcal{T}}(\sqrt{n^{K-1}} + \sqrt{n})$
$\frac{c\Lambda}{ S } \min_{j=1,\dots,K} \sqrt{r_j} B_T \left( \sqrt{n^{K-1}} + \sqrt{n} \right)$
$\frac{cA}{16!}\min_{i=1,\dots,K}\sqrt{\frac{r_i}{2}}B_{\mathcal{T}}(\sqrt{n^K}+n)$
$\frac{cA}{ \mathbf{S} (K-1)} \sum_{k=1}^{K-1} \min\left(\prod_{i=1}^{k} \sqrt{r_i}, \prod_{j=k+1}^{K} \sqrt{r_j}\right)$
$B_T \sqrt{n^{\lceil K/2 \rceil}}$
$\frac{cA}{ S } \left( \frac{\prod_{k=1}^{\kappa} r_k}{\max_{j=1,2} \prod_{i \in D_j} r_i} \right) \gamma_1(\mathcal{W}_{(D_1, D_2)})$
$\log(8)\left(\sqrt{n^{ D_1 }} + \sqrt{n^{ D_2 }}\right)$
$\frac{c\Lambda}{ S } \left(\frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\dots,K} r_j}\right) \gamma_1(\mathcal{W}) \log(4K) \sqrt{Kn}$
$\frac{c\Lambda}{ S }\left(\frac{\prod_{k=1}^{K}r_k}{\max_{j=1,\ldots,M}\prod_{i\in D_j}r_i}\right)\gamma_1(\mathcal{W}_{(D_1,\ldots,D_M)})$
$\log(4M)\sum_{i=1}^{M}\sqrt{n^{ D_j }}$
$\frac{c\Lambda}{ S } \min_{g \in G} \left( \frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\dots,M} \prod_{i \in D_j^{(S)}} r_i} \right)$
$\gamma_1(\mathcal{W}_{(D_1^{(g)},,D_{M_g}^{(g)})}^{(g)}) \max_{g \in G} \log(4M_g)$
$\sum_{j=1}^{M_g} \sqrt{n^{ M_g }}$

**Table 1** Rademacher complexities for convex norm regularized completion models for a *K*-mode tensor  $\mathcal{T} \in \mathbb{R}^{n \times \dots \times n}$  with a multilinear rank  $(r_1, \dots, r_K)$ .  $\gamma_1(\mathcal{X})$  is the largest singular value of  $\mathcal{X}$ , *G* reshaping sets of  $D^{(s)} = (D_1^{(s)}, \dots, D_{m_s}^{(s)})$ ,  $g = 1, \dots, G$ , and c, A, and  $B_T$  are constants

**Theorem 2** Let us consider a K-mode tensor  $\mathcal{W} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ . Let us consider a collection of G collection of reshaping sets  $D_L = (D^{(1)}, \dots, D^{(G)})$  where each  $D^{(s)} = (D_1^{(s)}, \dots, D_{M_s}^{(s)})$ ,  $s = 1, \dots, G$  consists a reshaping set for a  $M_s$ -mode reshaped tensor. Consider the hypothesis class  $W_{rl} = \{\mathcal{W}|\mathcal{W}^{(1)} + \dots + \mathcal{W}^{(G)} = \mathcal{W}, \|\mathcal{W}\|_{r\_latent(D_L)} \leq t\}$ for a given set of reshaping set  $(D_1, \dots, D_M)$ . Suppose that for all  $\mathcal{X}_{i_1, \dots, i_K}$ ,  $l(\mathcal{X}_{i_1, \dots, i_K}, \cdot)$  is  $\Lambda$ -Lipschitz continuous. Then,

(a) when W has a multilinear rank of  $(r_1, \ldots, r_K)$ , we obtain

$$\begin{split} R_{\mathrm{S}}(l \circ \mathcal{W}) &\leq \frac{cA}{|\mathrm{S}|} \min_{g \in G} \left( \frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\dots,M} \prod_{i \in D_j^{(g)}} r_i} \right) \gamma_1 \left( \mathcal{W}_{(D_1^{(g)},\dots,D_{M_g}^{(g)})}^{(g)} \right) \\ &\max_{g \in G} \log(4M_g) \sum_{j=1}^{M_g} \sqrt{\prod_{p \in D_j^{(g)}} n_p}. \end{split}$$

(b) when W has a CP rank of  $r_{cp}$ , we obtain

$$R_{S}(l \circ \mathcal{W}) \leq \frac{c\Lambda}{|S|} r_{cp} \min_{g} \gamma_1 \left( \mathcal{W}_{(D_1^{(g)}, \dots, D_{M_g}^{(g)})}^{(g)} \right)$$
$$\max_{g \in G} \log(4M_g) \sum_{j=1}^{M_g} \sqrt{\prod_{p \in D_j^{(g)}} n_p}.$$

where c is a constant.

Theorem 2 shows that latent reshaped tensor nuclear norm bounds the Rademacher complexity by the largest mode based component that results from all the reshaping sets. Further, with the multilinear rank of the tensor the Rademacher complexity is bounded by the smallest rank based component that results from all the reshaping sets. This observation indicates that properly selecting a set of reshaping sets to use with the latent reshaped tensor nuclear norm can lead to a lower generalization bound.

We want to point out that the largest singular values  $(\gamma_1(\cdot))$  that appear in both Theorems 1 and 2 can be upper bounded by taking the largest singular value with respect to all possible reshaping sets for a tensor. However, we do not use such a bounding to keep the Rademacher complexities small.

#### 4.1 Optimal reshaping strategies

Given that we have an understanding of the ranks of the tensor, Theorem 1 can be used to select a reshaping set such that reshaped tensor has a smaller rank and relatively smaller mode dimensions. However, since we do not know the rank in advance, selecting a reshaping set such that the reshaped tensor does not have large mode dimensions would lead to a better performance.

To avoid the difficulty in choosing a single reshaping set, we can use the reshaped latent tensor nuclear norm by choosing several reshaping sets that agree with our observation in Theorem 1. However, since the Rademacher complexity is bounded by the largest mode based components as shown in Theorem 2, it is important not to select reshaping sets that result in a tensor with large dimensions. A general strategy to create the reshaping sets by selecting the original tensor and other reshaping sets that do not result in large mode dimensions compared to the original tensor.

#### 5 Optimization procedures

It has been shown that learning by constraining the tensor nuclear norm is the NP-Complete problem (Hillar and Lim 2013), which makes solving the problems (3) and (4) computationally difficult. In Yang et al. (2015) an approximation method have been proposed to compute the spectral norm by computing largest singular vectors on each mode that is combined with the Frank-Wolfe optimization method to solve (3). We adopt their approximation method to solve our proposed completion models with reshaped tensors (3) and (4). We found that solutions using the approximation method provide agreements with our theoretical results related to generalization bounds we showed in the Sect. 4. However, there is no theoretical analysis available to understand how well the approximation method results in a solution compared to an exact solution.

The optimization method proposed in Yang et al. (2015) uses an approximation method for the spectral norm using a recursive algorithm based on singular value decomposition with respect to each mode. However, we adopt a more simpler approach as given in Algorithm 1, which we believe is more easier to implement. Using the approximation method, we provide an optimization procedure to solve the completion model that is regularized by a single reshaped norm in the Algorithm 2. The optimization procedure in Algorithm 2 is also similar to the Frank-Wolfe based optimization procedure proposed in Yang et al. (2015). The additions in Algorithm 2 to Yang et al. (2015) are the computation of the spectral norm of the reshaped tensor in step 7 and the conversion of the reshaped tensor to the original dimensions in step 12. Here, we want to recall Definition 1 to refer to the reshaping operator  $\Pi_{(D_1,...,D_M)}()$  and its inverse operator  $\Pi_{(D_1,...,D_M)}()$  for any given reshaping set  $(D_1, D_2, ..., D_M)$ .

1: Input: $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_K}$ 2: **Output:**  $w_1, ..., w_K, sv$ 3:  $\mathcal{Y} = \mathcal{A}$ 4: for k = 1, ..., K - 1 do  $M = \operatorname{reshape}(\mathcal{Y}, [n_k, n_{k+1} \cdots n_K])$ 5:6:  $(w_k, s_k, v_k) = \operatorname{svd}(M, 1)$ 7:  $\mathcal{Y} = \mathcal{Y} \times_1 w_k$  $\mathcal{Y} = \operatorname{reshape}(\mathcal{Y}, [n_{k+1}, \dots, n_K])$ 8: 9: end for  $\mathcal{Y} \times_1 \underline{w_{K-1}}$ 10:  $w_K = \frac{\mathbf{y}_{K-1}}{\|\mathbf{y}_{K-1}\|_2}$ 11:  $sv = \mathcal{A} \times_1 w_1 \times_2 w_2 \cdots \times_K w_K$ Algorithm 1: ApproxSpectralNorm( $\mathcal{A}$ )

- Input: X ∈ ℝ<sup>n<sub>1</sub>×n<sub>2</sub>···×n<sub>K</sub> with observed indexes Ω, Regularization parameter λ. Initial W<sup>0</sup>, Maximum iterations T, Reshaping dimensions (D<sub>1</sub>, D<sub>2</sub>,..., D<sub>M</sub>)
   Output: W<sup>T</sup>
  </sup>
- 3: t = 0
- 4: repeat
- 5: t = t + 1
- 6:  $f_{\mathcal{W}}(\mathcal{W}^t) = \frac{1}{2} \| \Omega(\mathcal{W}^t) \Omega(\mathcal{X}) \|_{\mathrm{F}}^2$

7: 
$$w_1, \ldots, w_K, sv = \text{ApproxSpectralNorm}(\Pi_{(D_1, \ldots, D_M)}(\nabla_{\mathcal{W}} f_{\mathcal{W}}(\mathcal{W}^t)))$$

- 8:  $\mathcal{W}_{descent}^t = -\lambda w_1 \otimes w_2, \dots \otimes w_K$
- 9: **if** linesearch == True **then**
- 10: Using an appropriate line search method (e.g. Yuan et al (2016))
- 11: else

12: 
$$\mathcal{W}^{t+1} = \mathcal{W}^t + \frac{2}{t+2} \Pi^{\top}_{(D_1,\dots,D_M)}(\mathcal{W}^t_{descent})$$

- 13: end if
- 14: **until** t = T

**Algorithm 2:** A Frank-Wolfe optimization method for a regularization with a reshaped tensor norm

1: Input:  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_K}$  with observed indexes  $\Omega$ , Regularization parameter  $\lambda$ . Initial  $\mathcal{W}^0$ , Maximum iterations T, G Reshaping sets  $(D^{(1)}, \ldots, D^{(G)})$  with  $D^{(g)} = (D_1^{(g)}, \dots, D_{M_q}^{(g)}), \ g = 1, \dots, G$ 2: Output:  $\mathcal{W}^T$ 3: t = 04: repeat  $5 \cdot$ t = t + 1 $f_{\mathcal{W}}(\mathcal{W}^t) = \frac{1}{2} \| \Omega(\mathcal{W}^t) - \Omega(\mathcal{X}) \|_{\mathrm{F}}^2$ 6:7: for  $g = 1, \ldots, G$  do  $w_1, \ldots, w_{M_g}, sv = \text{ApproxSpectralNorm}(\Pi_{(D_1^{(g)}, \ldots, D_{M_s}^{(g)})}(\nabla_{\mathcal{W}} f_{\mathcal{W}}(\mathcal{W}^t)))$ 8: 9:  $sv\_array(q) = sv$ 10: end for 11: $i = \arg\max_i(sv\_array)$ 12: $w_1, \ldots, w_{M_i}, sv = \text{ApproxSpectralNorm}(\Pi_{(D_1^{(i)}, \ldots, D_{M^{-}}^{(i)})}(\nabla_{\mathcal{W}} f_{\mathcal{W}}(\mathcal{W}^t)))$  $\mathcal{W}_{descent}^t = -\lambda \Pi_{(D_1^{(i)}, \dots, D_{M_i}^{(i)})}^{\top} (w_1 \otimes w_2, \dots \otimes w_{M_i})$ 13:14:if linesearch == True then Using an appropriate line search method (e.g. Yuan et al (2016)) 15:16:else  $\mathcal{W}^{t+1} = \mathcal{W}^t + \frac{2}{t+2}(\mathcal{W}^t_{descent})$ 17:18:end if 19: **until** t = T

**Algorithm 3:** A Frank-Wolfe optimization method for a regularization with the reshaped latent tensor nuclear norm

Next, we give an algorithm to solve the completion model regularized by the reshaped latent tensor nuclear norm. The Frank-Wolfe optimization method has also been applied to efficiently solve learning models regularized by the latent trace norms (Guo et al. 2017). We follow their approach to design Frank-Wolfe method for the reshaped latent tensor nuclear norm and Algorithm 3 shows the steps for optimization. From Lemma 1, we know that we need to find the reshaping with the largest spectral norm each t step to update the Frank-Wolfe procedure. This is shown in the lines 7–11 in the Algorithm 3.

# 6 Experiments

In this section, we give simulation and real-data experiments.

## 6.1 Simulation experiments

We created simulation experiments for tensor completion using tensors with some fixed multilinear rank and CP rank. We create a *K*-mode tensor with the multilinear rank of  $(r_1, \dots, r_K)$  by generating a tensor  $\mathcal{T} \in \mathbb{R}^{n_1 \times \dots \times n_K}$  using the Tucker decomposition (Kolda and Bader 2009) as  $\mathcal{T} = \mathcal{C} \times_1 U_1 \times_2 U_2 \times_3 \dots \times_K U_K$ , where  $\mathcal{C} \in \mathbb{R}^{r_1 \times \dots \times r_K}$  is a core tensor whose elements are sampled from a normal distribution specifying the multilinear rank  $(r_1, \dots, r_K)$  and  $U_k \in \mathbb{R}^{r_k \times n_k}$ ,  $k = 1, \dots, K$  are orthogonal component matrices. We create a tensor with the CP rank of *r* using the CP decomposition (Kolda and Bader 2009) as  $\mathcal{T} = \sum_{i=1}^r c_i u_{1i} \otimes u_{2i} \otimes \dots \otimes u_{Ki}$  where  $u_{ki} \in \mathbb{R}^{n_k}$ ,  $k = 1, \dots, K$ ,  $i = 1, \dots, r$  are sampled

from a normal distribution and normalized such that  $||u_{ki}||_2^2 = 1$  and  $c_i \in \mathbb{R}^+$ . From the total number of elements in the generated tensors, we randomly selected 10, 40, and 70 percentages as training sets, and from the remaining we selected 10 percent of elements as validation set, and the rest were taken as test data. We conducted 3 simulations for each randomly generated tensor.

For all simulation experiments, we tested completion using our proposed completion models (3) with the reshaped tensor nuclear norm (abbreviated as RTNN) and (4) with the reshaped latent tensor nuclear norm (abbreviated as RLTNN). Additionally, we performed completion using the tensor nuclear norm (abbreviated as TNN) without reshaping and the square norm (abbreviated as SN). As further baseline methods, we used tensor completion with regularization using the overlapped trace norm (abbreviated as OTN), scaled latent trace norm (abbreviated as SLTN), and the Schatten TT norm (Imaizumi et al. 2017) (abbreviated as STTN). As the performance measure of completion, we calculated the mean squared error (MSE) on the validation data and test data. For all completion models, we performed cross-validation of regularization parameters in power of  $2^x$ , with *x* ranging from -5 to 15 with intervals of 0.25.

For our first simulation experiment, we created a 4-way tensors  $T_1 \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times n_4}$  with  $n_1 = n_2 = 10, n_3 = n_4 = 40$  with a multilinear rank of  $(r_1, r_2, r_3, r_4) = (9, 9, 3, 3)$ . From Mu et al. (2014) we can reshape  $T_1$  by using a reshaping set of  $(D_1, D_2) = ((n_1, n_3), (n_2, n_4))$  such that it creates a square matrix for the square norm. From Theorem 1, we see that the rank components in the Rademacher complexity for the nuclear norm and the square norm are  $\prod_{k=1}^{K} r_k/(\max_{j=1,...,4} r_j) = 243$  and  $\prod_{k=1}^{K} r_k/(\max_{j=1,2} \prod_{i \in D_j} r_i) = 27$ , respectively. Further, Theorem 1 shows that the mode based components for the nuclear norm and the square norm are  $\log(4 \cdot 4)(\sum_{k=1}^4 \sqrt{n_k}) \approx 53$  and  $\log(4 \cdot 2)(\sqrt{n_1n_3} + \sqrt{n_2n_4}) \approx 83$ , respectively. This leads to a lower generalization bound for the nuclear norm compared to the square norm justifying its better performance as shown in the Fig. 1a. However, Theorem 1 indicates that the lowest generalization bound is obtained by using the reshaping set  $(D'_1, D'_2, D'_3) = ((n_1, n_2), n_3, n_4)$ , which combines the high ranked modes (mode 1 and mode 2) together resulting in a rank based component of  $\prod_{k=1}^{K} r_k/(\max_{j=1,2,3} \prod_{i \in D'_j} r_i) = 9$  and a mode based component of  $\log(4 \cdot 3)(\sqrt{n_1n_2} + \sqrt{n_3} + \sqrt{n_4}) \approx 56$ . Figure 1a agrees with our theoretical analysis showing that our proposed reshaped tensor nuclear norm obtains



**Fig. 1** Performances of completion of the tensors (**a**) Tensor  $\mathcal{T}_1 \in \mathbb{R}^{10 \times 10 \times 40 \times 40}$  with a multilinear rank (9, 9, 3, 3) and (**b**) tensor  $\mathcal{T}_2 \in \mathbb{R}^{10 \times 10 \times 10 \times 10 \times 10}$  with a CP rank 3

519

the best performance compared to other norms. For the reshaped latent tensor nuclear norm, we combined the two reshaping sets  $((n_1, n_2, n_3, n_4), ((n_1, n_2), n_3, n_4))$ . Applying Theorem 2, we see that this reshaping set combination leads to a lower Rademacher complexity. However, this combination only gave a comparable performance compared to the reshaped tensor nuclear norm.

As our second simulation, we created a 5-mode tensor  $T_2 \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_5}$ , where  $n_1 = n_2 = \dots = n_5 = 10$  with a CP rank of 3. From Theorem 1, we know that we can only consider the mode based component of the Rademacher complexity to obtain a lower generalization bound. For the square norm we can use a reshaping set such as  $(D_1, D_2) = ((n_1, n_2), (n_3, n_4, n_5))$ , which results in the mode based component as  $\log(4 \cdot 2)(\sqrt{n_1n_2} + \sqrt{n_3n_4n_5}) \approx 86$ . The tensor nuclear norm leads to a mode based component of  $\log(4 \cdot 5)(\sum_{k=1}^{5} \sqrt{n_k}) \approx 47$ . As an alternative reshaping method, we propose to combine any two modes together to create a reshaping set such as  $(D'_1, D'_2, D'_2) = (n_1, n_2, n_3, (n_4, n_5))$  for the reshaped tensor nuclear norm, which lead to a mode based component of  $\log(4 \cdot 4)(\sqrt{n_1} + \sqrt{n_2} + \sqrt{n_3} + \sqrt{n_4n_5}) \approx 54$ . Comparing the Rademacher complexities using the mode based components we see that the lowest generalization bound is given by the tensor nuclear norm. Figure 1b shows that our theoretical observation is accurate since the tensor nuclear norm gives the best performance compared to other two reshaped norms. For the reshaped latent tensor nuclear norm we used all the 10 combinations of two modes combinations, which resulted in reshaping sets of  $D = (((n_1, n_2), n_3, n_4, n_5), (n_1, (n_2, n_3), n_4, n_5), \dots, (n_1, n_2, n_3, (n_4, n_5))))$ . Figure 1b shows that the reshaped latent tensor nuclear norm has outperformed the tensor nuclear norm.

The next simulation focuses on a different multilinear rank for the 4-way tensor  $\mathcal{T}_3 \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times n_4}$  with  $n_1 = n_2 = 10, n_3 = n_4 = 40$ . Figure 2a shows the simulation experiment with multilinear rank of (3, 3, 35, 35). Again based on Mu et al. (2014) we can  $T_3$  by using a reshaping set of  $(D_1, D_2) = ((n_1, n_3), (n_2, n_4))$  or reshape  $(D_1, D_2) = ((n_1, n_4), (n_2, n_3))$  to create a square matrix to use with the square norm. From Theorem 1 we can observe that the square norm will result in a rank based component of  $\prod_{k=1}^{K} r_k / (\max_{j=1,2} \prod_{i \in D_j} r_i) = 105$ and а mode based component of  $\log(4 \cdot 2)(\sqrt{n_1 n_3} + \sqrt{n_2 n_4}) \approx 63$ . However, if we combine the high ranked modes 3 and 4 together to create the reshaping set  $(D'_1, D'_2, D'_3) = (n_1, n_2, (n_3, n_4))$  for the reshaped tensor based nuclear norm, then the rank component will decrease to



**Fig. 2** Performances of completion of the tensors (**a**) Tensor  $\mathcal{T}_3 \in \mathbb{R}^{10 \times 10 \times 40 \times 40}$  with a multilinear rank (3, 3, 35, 35) and (**b**) tensor  $\mathcal{T}_3 \in \mathbb{R}^{10 \times 10 \times 10 \times 10 \times 10}$  with a CP rank 243

 $\prod_{k=1}^{K} r_k / (\max_{j=1,2,3} \prod_{i \in D'_j} r_i) = 9 \text{ and mode based component will decrease to } \log(4 \cdot 2)(\sqrt{n_1}\sqrt{n_3} + \sqrt{n_2n_4}) \approx 55.$  Furthermore, the tensor nuclear norm leads to a rank based component of  $\prod_{k=1}^{K} r_k / (\max_{j=1,\dots,4} r_j) = 315$  and mode based component of  $\log(4 \cdot 4)(\sum_{k=1}^{4} \sqrt{n_k}) \approx 53$  resulting in a larger generalization bound compared to the proposed reshaped set  $(D'_1, D'_2, D'_3) = (n_1, n_2, (n_3, n_4))$ . This analysis is also confirmed with the experimental results as shown in Fig. 2a where the reshaped tensor nuclear norm gives the best performance. Using the Theorem 2, we find that if we use reshaping sets  $((n_1, n_2, n_3, n_4), ((n_1, n_2), n_3, n_4))$  for the reshaped latent tensor nuclear norm, the Rademacher complexity will be bounded by the smaller rank based component from the reshaping set  $((n_1, n_2), n_3, n_4)$ . However, the reshaped latent tensor nuclear norm was not able to perform better than the tensor nuclear norm or the proposed reshaped norm with  $(D'_1, D'_2, D'_3) = (n_1, n_2, (n_3, n_4))$ .

The final simulation result shown in Fig. 2b is for a tensor  $\mathcal{T}_4 \in \mathbb{R}^{10 \times 10 \times 10 \times 10 \times 10}$  with a CP rank of 243. For this experiment, we used the same reshaping strategies as in the previous experiment with CP rank. We see that when the fraction of training samples is less than 40 percent the tensor nuclear norm has given the best performance. When the fraction of the training samples increases beyond 40 percent the reshaped latent tensor nuclear norm has outperformed the tensor nuclear norm.

#### 6.2 Multi-view video completion

We performed completion on multi-view video data using the EPFL data set: Multi-camera Pedestrian Videos data (Berclaz et al. 2011). Videos in this data set capture sequentially entering a room and walking around of four people from 4 views using 4 synchronized cameras. We down-sampled each video frame to a height of 96 and width of 120 to obtain a frame as a RGB-color image with dimensions of  $96 \times 120 \times 3$ . We sequentially selected 391 frames from each video. Combining all the video frames from all views resulted in a tensor of dimensions of  $96 \times 120 \times 3 \times 391 \times 4$  (height × width × color × frames × views).

To evaluate completion, we randomly removed entries from the multi-view tensor and performed completion using the remaining elements. We randomly selected percentages of 2, 4, 8, 16, 32, and 64 of the total number of elements in the tensor as training elements. As our validation set we selected 10 percent of the total number of elements. The rest of the remaining elements were taken as the test set. To create a square norm, we considered the reshaping set ((height,width),(color,frames,views)). For the reshaped tensor nuclear norm, we experimentally found that the reshaping set ((height,views),(width,color),(frames)) gives the best performance. To create the reshaping set for the reshaped tensor nuclear norm we combined the reshaping sets of the square norm and the reshaped tensor nuclear norm with the unreshaped original tensor. The resulting set was D = ((height,width,color,frames,views), ((height,width), (color,frames,views)),((height,views),(width,color),(frames))). We cross-validated all the completion models $with regularization parameters out of <math>10^{-1}$ ,  $10^{-0.75}$ ,  $10^{-0.5}$ , ...,  $10^7$ .

Figure 3 shows that when the training set is small (or the reshaped tensor nuclear norm is sparse) the reshaped tensor nuclear norm and the tensor nuclear norm have given good performance compared to the square norm. When the percentage of observed elements increases more than 16 percent, the square norm outperforms the other norms. However,





the reshaped latent tensor nuclear norm has shown to be adaptive to all fractions of training samples and has given the overall best performance.

## 7 Conclusions

In this paper, we generalize tensor reshaping for low-rank tensor regularization and introduce the reshaped tensor nuclear norm and the reshaped latent tensor nuclear norm. We propose tensor completion models that are regularized by the proposed norm. Using generalization bound analysis of the proposed completion models we show that the proposed norms lead to smaller Rademacher complexity bounds compared to exiting norms. Further, using our theoretical analysis we discuss optimal conditions to create reshaped tensor nuclear norms. Simulation and real-data experiments confirm our theoretical analysis.

Our research opens up several future research directions. The most important research should be focused on developing theoretical guaranteed methods for optimization of completion models regularized by the the proposed tensor nuclear norms. Though the approximation methods we have adopted for computing the tensor spectral norm to be used with the Frank-Wolfe from Yang et al. (2015) provide performances that agrees with our generalization bounds we do not know its approximation error. We believe that future theoretical investigations are needed to understand qualitative properties of the proposed optimization procedures. Furthermore, optimization methods for nuclear norms that can scale for large-scale higher order tensors would be an important future research direction. Another important research direction is to further explore the theoretical foundation of tensor completion using the reshaped tensor nuclear norm. In this regard, recovery bounds (Yuan and Zhang 2016) would provide us with stronger bounds on sample complexities for the proposed method.

Acknowledgements H.M. has been supported in part by JST ACCEL [Grant Number JPMJAC1503], MEXT Kakenhi [Grant Number 19H04169] and AIPSE by Academy of Finland.

# Appendix

#### Dual norms of reshaped tensor nuclear norms

In this section, we discuss the dual norm of the proposed reshaped tensor nuclear norm. The dual norm is useful in developing optimization procedures and proving theoretical bounds.

The dual norm of the tensor nuclear norm (Yang et al. 2015; Yuan and Zhang 2016) for a *K*-mode tensor  $T \in \mathbb{R}^{n_1 \times \cdots \times n_K}$  is given by

$$\|\mathcal{T}\|_{\mathrm{op}} = \max_{\|y_i\|_2 = 1, 1 \le i \le K} \langle \mathcal{T}, y_1 \otimes y_2 \otimes \cdots \otimes y_K \rangle.$$
(6)

This definition applies to all tensor nuclear norms including the reshaped norms.

The next lemma provides the dual norm for the reshaped latent tensor nuclear norm.

**Lemma 1** The dual norm of the reshaped latent tensor nuclear norm for a tensor  $\mathcal{W} \in \mathbb{R}^{n_1 \times \cdots \times n_k}$  for a collection of G reshaping sets  $D_L = (D^{(1)}, \dots, D^{(G)})$  is

$$\|\mathcal{W}\|_{r\_\text{latent}(D_{\mathsf{L}})^*} = \max_g \|\mathcal{W}_{(D^{(g)})}\|_{\text{op}}.$$

**Proof** Using the standard formulation of the dual norm, we write the dual norm for  $\|\mathcal{W}\|_{r | latent(D_r)^*}$  as

$$\|\mathcal{W}\|_{r_{latent}(D_{L})^{*}} = \sup\left\langle \sum_{k=1}^{G} \mathcal{X}^{(k)}, \mathcal{W} \right\rangle \quad \text{s.t.} \quad \inf_{\mathcal{X}^{(1)} + \dots + \mathcal{X}^{(G)} = \mathcal{X}} \sum_{k=1}^{G} \|\mathcal{X}^{(k)}_{(D^{(k)})}\|_{\star} \le 1.$$
(7)

The solution to (7) resides on the simplex of  $\inf_{\mathcal{X}^{(1)}+\dots+\mathcal{X}^{(G)}=\mathcal{X}} \sum_{k=1}^{G} \|\mathcal{X}_{(D^{(k)})}^{(k)}\|_{\star} \leq 1$  and one of the edges of the simplex is a solution. Then, we can take any  $g \in 1, \dots, G$  such that  $\mathcal{X}^{(g)} = \mathcal{X}$  and all  $\mathcal{X}^{(k\neq g)} = 0$ , and arrange (7) as

$$\|\mathcal{W}\|_{\mathrm{r\_latent}(D_{\mathrm{L}})^*} = \sup_{g \in 1, \dots, G} \left\langle \mathcal{X}_{(D^{(g)})}, \mathcal{W}_{(D^{(g)})} \right\rangle \quad \text{s.t. } \|\mathcal{X}_{(D^{(g)})}\|_{\star} \le 1,$$

which results in the following

$$\|\mathcal{W}\|_{\mathrm{r\_latent}(D_{\mathrm{L}})^{*}} = \max_{g \in 1, \dots, G} \|\mathcal{W}_{(D^{(g)})}\|_{\mathrm{op}}.$$

#### Proofs of theoretical analysis

In this section, we provide proofs of the theoretical analysis in Sect. 4.

First, we prove following useful lemmas. These lemmas bound the tensor nuclear norm and the reshaped tensor nuclear norms with respect to the multilinear rank of a tensor.

**Lemma 2** Let  $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$  be a random K-mode tensor with a multilinear rank of  $(r_1, \ldots, r_K)$ . Let  $r_{cp}$  be the CP rank of  $\mathcal{X}$ , then

$$\begin{split} \|\mathcal{X}\|_{\star} &= \left\{ \sum_{j=1}^{r_{cp}} \gamma_j |\mathcal{X} = \sum_{j=1}^{r_{cp}} \gamma_j u_{1j} \otimes u_{2j} \otimes \cdots \otimes u_{Kj}, \|u_{kj}\|_2^2 = 1, \gamma_j \ge \gamma_{j+1} > 0 \right\} \\ &\leq \frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\dots,K} r_i} \gamma_1, \end{split}$$

where  $\gamma_i$  is the *i*th singular value of  $\mathcal{X}$ .

**Proof** Let us consider the Tucker decomposition of  $\mathcal{X}$  as

$$\mathcal{X} = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \cdots \sum_{j_K=1}^{r_K} \mathcal{C}_{j_1,\dots,j_K} u_{j_1}^{(1)} \otimes u_{j_2}^{(2)} \otimes \cdots \otimes u_{j_K}^{(K)},$$

where  $C \in \mathbb{R}^{r_1 \times \cdots \times r_K}$  is the core tensor and  $u_{(i)}^j \in \mathbb{R}^{n_j}$ ,  $\|u_{(i)}^j\|_2 = 1$ ,  $i = 1, \dots, r_i$ ,  $j = 1, \dots, K$  are component vectors.

Following Chapter 8 of Hackbusch (2012), we can express the above Tucker decomposition as

$$\mathcal{X} = \sum_{j_2=1}^{r_2} \cdots \sum_{j_K=1}^{r_K} \underbrace{\left(\sum_{j_1=1}^{r_1} \mathcal{C}_{j_1,\dots,j_K} u_{j_1}^{(1)}\right)}_{\hat{u}^{(1)}[j_2,\dots,j_K] \in \mathbb{R}^{n_1}} \otimes u_{j_2}^{(2)} \otimes \cdots \otimes u_{j_K}^{(K)},\tag{8}$$

where we have taken summation over the multiplications of core tensor elements and component vectors of the mode 1. It is easy to see that we can also consider the summation over component vectors of any other mode in a similar manner.

By considering  $\hat{u}^{(1)}[j_2, \dots, j_K] = \gamma[j_2, \dots, j_K] \frac{\hat{u}^{(1)}[j_2, \dots, j_K]}{\|\hat{u}^{(1)}[j_2, \dots, j_K]\|_2}$  where  $\gamma[j_2, \dots, j_K] = \|\hat{u}^{(1)}[j_2, \dots, j_K]\|_2$ , the above arrangement leads to a CP decomposition with a rank of  $r_{cp} = \frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\dots,K} r_j}$ .

By arranging  $\gamma[j_2, ..., j_K]$  in descending order along component vectors  $\hat{u}^{(1)}[j_2, ..., j_K]$ and renaming them as  $\gamma_1 \ge \gamma_2 \ge ...$  and  $u_{1j}$ , respectively, we obtain

$$\|\mathcal{X}\|_{\star} = \bigg\{ \sum_{j=1}^{r_{cp}} \gamma_j | \mathcal{X} = \sum_{j=1}^{r_{cp}} \gamma_j u_{1j} \otimes u_{2j} \cdots \otimes u_{Kj}, \|u_{kj}\|_2^2 = 1, \gamma_j \ge \gamma_{j+1} > 0 \bigg\},\$$

where  $u_{kj} \in [u_1^{(k)}, \dots, u_{r_k}^{(k)}]$  are component vectors from (8) for each  $k = 2, \dots, K$ . Then we arrive at the final bound of

Then we arrive at the final bound of

$$\begin{split} \|\mathcal{X}\|_{\star} &= \left\{ \sum_{j=1}^{r_{cp}} \gamma_{j} | \mathcal{X} = \sum_{j=1}^{r_{cp}} \gamma_{j} u_{1j} \otimes u_{2j} \cdots \otimes u_{Kj}, \|u_{kj}\|_{2}^{2} = 1, \gamma_{j} \geq \gamma_{j+1} > 0 \right\} \\ &\leq \frac{\prod_{k=1}^{K} r_{k}}{\max_{j=1,\dots,K} r_{i}} \gamma_{1}. \end{split}$$

**Lemma 3** Let  $\mathcal{X} \in \mathbb{R}^{n \times ... \times n}$  be a random K-mode tensor with multilinear rank of  $(r_1, ..., r_K)$ . We consider a set of M reshaping modes  $D_i$ , i = 1, ..., M. Let  $r_{cp}$  be the CP rank of  $\mathcal{X}$ , then

$$\begin{split} \|\mathcal{X}_{(D_1,\dots,D_M)}\|_{\star} &= \left\{ \sum_{j=1}^{r_{cp}} \gamma_j |\mathcal{X}_{(D_1,\dots,D_M)} = \sum_{j=1}^{r_{cp}} \gamma_j u_{1j} \otimes u_{2j} \cdots \otimes u_{Mj}, \\ \|u_{kj}\|_2^2 &= 1, \gamma_j \ge \gamma_{j+1} > 0 \right\} \le \frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\dots,M} \prod_{i \in D_j} r_i} \gamma_1, \end{split}$$

where  $\gamma_i$  is the *i*th singular value of  $\mathcal{X}_{(D_1,...,D_M)}$ .

**Proof** Let us consider the Tucker decomposition of  $\mathcal{X}$  as

$$\mathcal{X} = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \cdots \sum_{j_K=1}^{r_K} \mathcal{C}_{j_1, \dots, j_K} u_{j_1}^{(1)} \otimes u_{j_2}^{(2)} \otimes \cdots \otimes u_{j_K}^{(K)},$$

where  $C \in \mathbb{R}^{r_1 \times \cdots \times r_K}$  is the core tensor and  $u_i^j \in \mathbb{R}^{n_j}$ ,  $||u_i^j||_2 = 1$ ,  $i = 1, \dots, r_i, j = 1, \dots, K$ are component vectors. We rearrange the Tucker decomposition for the reshaped tensor  $\mathcal{X}_{(D_1,\dots,D_M)}$  as

$$\begin{split} \mathcal{X}_{(D_1,\ldots,D_M)} &= \sum_{j'_{a'},j'_{b'},\ldots\in D_2} \cdots \left(\sum_{\substack{\sum_{j'_{a''},j'_{b''},\ldots\in D_M}\\ j_{a''},j'_{b''},\ldots\in D_M}} \underbrace{\left(\sum_{j_a,j_b,\ldots\in D_1} \mathcal{C}_{j_1,\ldots,j_K} \Pi_{D_1}(u_{j_a}^{(a)}\otimes u_{j_b}^{(b)}\cdots)\right)}_{\hat{u}_1[D'_2,\ldots,D'_M]\in \mathbb{R}^{\mathrm{prod}(D_1)}} \\ &\otimes \Pi_{D_M}(u_{j'_a}^{(a')}\otimes u_{j'_b}^{(b')}\cdots)\right)\otimes \cdots, \end{split}$$

with  $\hat{u}_1[D'_2, \dots, D'_M] = \gamma[D'_2, \dots, D'_M] \frac{\hat{u}_1[D'_2, \dots, D'_M]}{\|\hat{u}_1[D'_2, \dots, D'_M]\|_2}$  where  $\gamma[D'_2, \dots, D'_M] = \|\hat{u}_1[D'_2, \dots, D'_M]\|_2$ . We can consider the above summation over any reshaping set and it is easy to see that the arrangement takes a CP decomposition with a CP rank of  $r_{cp} = \frac{\prod_{k=1}^K r_k}{\max_{j=1,\dots,M} \prod_{i \in D_j} r_i}$ .

By arranging  $\gamma[D_2, ..., D_M]$  in descending order order along with component vectors  $\hat{u}^{(1)}[D_2, ..., D_M]$  and renaming them as  $\gamma_1 \ge \gamma_2 \ge ...$  and  $u_{1j}$ , respectively, we obtain

$$\|\mathcal{X}_{(D_1,\ldots,D_M)}\|_{\star} = \left\{ \sum_{j=1}^{r_{cp}} \gamma_j |\mathcal{X}_{(D_1,\ldots,D_M)} = \sum_{j=1}^{r_{cp}} \gamma_j u_{1j} \otimes u_{2j} \cdots \otimes u_{Mj}, \|u_{kj}\|_2^2 = 1, \gamma_j \ge \gamma_{j+1} > 0 \right\},$$

where  $u_{kj} \in [\Pi_{D_k}(u_1^{(a')} \otimes u_1^{(b')} \cdots), \dots, \Pi_{D_k}(u_{r'_a}^{(a')} \otimes u_{r'_b}^{(b')} \cdots)]$  are components for each  $k = 2, \dots, M$  and  $a', b', \dots \in D_k$ .

Using the above results we arrive at the following bound

$$\begin{aligned} \|\mathcal{X}_{(D_{1},...,D_{M})}\|_{\star} &= \left\{ \sum_{j=1}^{r_{cp}} \gamma_{j} | \mathcal{X}_{(D_{1},...,D_{M})} = \sum_{j=1}^{r_{cp}} \gamma_{j} u_{1j} \otimes u_{2j} \cdots \otimes u_{Mj}, \\ \|u_{kj}\|_{2}^{2} &= 1, \gamma_{j} \geq \gamma_{j+1} > 0 \right\} \leq \frac{\prod_{k=1}^{K} r_{k}}{\max_{j=1,...,M} \prod_{i \in D_{j}} r_{i}} \gamma_{1}, \end{aligned}$$

**Lemma 4** Let  $\mathcal{X} \in \mathbb{R}^{n_1 \times \ldots \times n_K}$  be a random K-mode tensor with CP rank of  $r_{cp}$ . We consider a set of M reshaping sets  $D_i$ ,  $i = 1, \ldots, M$ . Then

$$\|\mathcal{X}_{(D_1,\ldots,D_M)}\|_{\star} \leq r_{cp}\gamma_1,$$

where  $\gamma_i$  is the *i*th singular value of  $\mathcal{X}_{(D_1,...,D_M)}$ .

**Proof** Let us consider  $\mathcal{X}$  as

$$\mathcal{X} = \sum_{j=1}^{r_{cp}} \gamma_j u_{1j} \otimes u_{2j} \cdots \otimes u_{Kj},$$

with  $\|u_{kj}\|_2^2 = 1, \gamma_j \ge \gamma_{j+1} > 0$ . For the reshaping set  $(D_1, \dots, D_M)$ , we rearrange  $\mathcal{X}$  as

$$\mathcal{X}_{(D_1,...,D_M)} = \sum_{j=1}^{r_{cp}} \gamma_j(\circ_{i_1 \in D_1} u_{i_1 j}) \otimes (\circ_{i_2 \in D_2} u_{i_2 j}) \cdots \otimes (\circ_{i_M \in D_M} u_{i_M j}),$$

where  $a \circ b = [a_1 b, a_2 b, ..., a_n b]^{\top}$  is the Khatri-Rao product (Kolda and Bader 2009). It is easy to verify that  $\operatorname{vec}((a \circ b) \otimes (c \circ d)) = \operatorname{vec}(a \otimes b \otimes c \otimes d)$ , which indicates that  $\operatorname{vec}(\mathcal{X}) = \operatorname{vec}(\mathcal{X}_{(D_1,...,D_M)})$ .

Using the fact that  $\operatorname{Rank}(a \otimes b) \leq \operatorname{Rank}(a)\operatorname{Rank}(b)$  from Kolda and Bader (2009), we have

$$\operatorname{Rank}(\mathcal{X}_{(D_1,\ldots,D_M)}) \leq \operatorname{Rank}(\mathcal{X}) = r_{cp}.$$

This lead to the final observation

$$\|\mathcal{X}_{(D_1,\ldots,D_M)}\|_{\star} \leq r_{cp}\gamma_1.$$

In order to prove Rademacher complexities in Theorems 1 and 2, we use the following lemma form Raskutti et al. (2015).

**Lemma 5** (Raskutti et al. 2015) Consider a K-mode tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$  with random samples from an i.i.d. Gaussian tensor ensemble. Then

$$\mathbb{E} \|\mathcal{X}\|_{\text{op}} \le 4 \log(4K) \sum_{k=1}^{K} \sqrt{n_k}.$$

Given a tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$  with Gaussian entries, we can write

$$\mathbb{E}\mathcal{X} = \mathbb{E}\sum_{i_1,i_2,\ldots,i_K}\mathcal{X}_{i_1,i_2,\ldots,i_K}e_{i_1}\otimes e_{i_2}\otimes\cdots\otimes e_{i_K},$$

where  $e_{i_k}$  is the vector with 1 at the kth element and rest of the elements are zero. Due to each  $\mathcal{X}_{i_1,i_2,\ldots,i_k}^{\kappa}$  being a Gaussian entry, we have

$$\mathbb{E}\mathcal{X} = \mathbb{E}_{g}\mathbb{E}_{\epsilon}\sum_{i_{1},i_{2},\ldots,i_{K}}\epsilon_{i_{1},i_{2},\ldots,i_{K}}|\mathcal{X}_{i_{1},i_{2},\ldots,i_{K}}|e_{i_{1}}\otimes e_{i_{2}}\otimes\cdots\otimes e_{i_{K}},$$

where  $\epsilon_{i_1,i_2,\ldots,i_k} \in \{-1,1\}$ . Using the Jensen's inequality, we have

$$\begin{split} \mathbb{E}_{g} \mathbb{E}_{\epsilon} \sum_{i_{1}, i_{2}, \dots, i_{K}} \epsilon_{i_{1}, i_{2}, \dots, i_{K}} | \mathcal{X}_{i_{1}, i_{2}, \dots, i_{K}} | e_{i_{1}} \otimes e_{i_{2}} \otimes \cdots e_{i_{K}} \\ \geq \mathbb{E}_{\epsilon} \sum_{i_{1}, i_{2}, \dots, i_{K}} \epsilon_{i_{1}, i_{2}, \dots, i_{K}} \mathbb{E}_{g} | \mathcal{X}_{i_{1}, i_{2}, \dots, i_{K}} | e_{i_{1}} \otimes e_{i_{2}} \otimes \cdots e_{i_{K}} \\ \geq \sqrt{2\pi} \mathbb{E}_{\epsilon} \sum_{i_{1}, i_{2}, \dots, i_{K}} \epsilon_{i_{1}, i_{2}, \dots, i_{K}} e_{i_{1}} \otimes e_{i_{2}} \otimes \cdots e_{i_{K}}. \end{split}$$

This shows that we can use the Lemma 3 to bound tensors with Bernoulli random variables.

Next we give the detailed proof of Theorem 1.

**Proof of Theorem 1** We expand the Rademacher complexity in (5) as

$$R_{\mathbf{S}}(l \circ \mathcal{W}) = \frac{1}{|\mathbf{S}|} \mathbb{E}_{\sigma} \left[ \sup_{\mathcal{W} \in W} \sum_{i_1, \dots, i_K} \Sigma_{i_1, \dots, i_K} l \left( \mathcal{X}_{i_1, \dots, i_K}, \mathcal{W}_{i_1, \dots, i_K} \right) \right],$$

where  $\Sigma_{i_1,...,i_K} = \sigma_j$  when  $(i_1, ..., i_K) \in S$  and  $\Sigma_{i_1,...,i_K} = 0$ , otherwise. We analyze the Rademacher complexity

$$R_{\rm S}(l \circ \mathcal{W}) = \frac{1}{|{\rm S}|} \mathbb{E}_{\sigma} \left[ \sup_{\mathcal{W} \in W} \sum_{i_1, \dots, i_K} \Sigma_{i_1, \dots, i_K} l(\mathcal{X}_{i_1, \dots, i_K}, \mathcal{W}_{i_1, \dots, i_K}) \right],$$
  

$$\leq \frac{\Lambda}{|{\rm S}|} \mathbb{E}_{\sigma} \left[ \sup_{\mathcal{W} \in W} \sum_{i_1, \dots, i_K} \Sigma_{i_1, \dots, i_K} \mathcal{W}_{i_1, \dots, i_K} \right], (\text{Rademacher contraction}) \qquad (9)$$
  

$$\leq \frac{\Lambda}{|{\rm S}|} \mathbb{E}_{\sigma} \sup_{\mathcal{W} \in W} \left\| \mathcal{W}_{(D_1, \dots, D_M)} \right\|_{\star} \left\| \Sigma_{(D_1, \dots, D_M)} \right\|_{\star^*}, (\text{Duality relationship})$$

(a) Given that tensor has a multilinear rank of  $(r_1, \ldots, r_K)$ , using the Lemma 3 we know that

$$\|\mathcal{W}_{(D_1,...,D_M)}\|_{\star} \leq \left(\frac{\prod_{k=1}^K r_k}{\max_{j=1,...,M} \prod_{i \in D_j} r_i}\right) \gamma_1(\mathcal{W}_{(D_1,...,D_M)}).$$
 (10)

Using Lemma 5 we can bound  $\mathbb{E}_{\sigma} \left\| \Sigma_{(D_1, \dots, D_M)} \right\|_{\star *}$  as

$$\mathbb{E}_{\sigma} \left\| \Sigma_{\left(D_{1},\ldots,D_{M}\right)} \right\|_{\star^{*}} \leq 4 \log\left(4M\right) \sum_{j=1}^{M} \sqrt{\prod_{p \in D_{j}} n_{p}}.$$
(11)

By substituting (10) and (11) to (9), we obtain the following bound

## Deringer

$$R_{\rm S}(l\circ\mathcal{W}) \le \frac{c\Lambda}{|{\rm S}|} \left(\frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\dots,M} \prod_{i\in D_j} r_i}\right) \gamma_1(\mathcal{W}_{(D_1,\dots,D_M)}) \log(4M) \sum_{j=1}^{M} \sqrt{\prod_{p\in D_j} n_p}.$$
 (12)

(b) Given that a tensor has a CP rank of  $r_{cp}$ , using the Lemma 4 we have

$$\|\mathcal{W}_{(D_1,...,D_M)}\|_{\star} \le r_{cp}\gamma_1(\mathcal{W}_{(D_1,...,D_M)}).$$
(13)

From Lemma 5, we have

$$\mathbb{E}_{\sigma} \left\| \Sigma_{\left(D_{1},\dots,D_{M}\right)} \right\|_{\star^{*}} \leq 4 \log\left(4M\right) \sum_{j=1}^{M} \sqrt{\prod_{p \in D_{j}} n_{p}}.$$
(14)

By substituting (13) and (14) to (9), we obtain the desired bound

$$R_{\mathrm{S}}(l\circ\mathcal{W}) \leq \frac{c\Lambda}{|\mathrm{S}|} r_{cp} \gamma_1(\mathcal{W}_{(D_1,\dots,D_M)}) \log(4M) \sum_{j=1}^M \sqrt{\prod_{p \in D_j} n^{|D_j|}}.$$
 (15)

Next, we give the proof for Theorem 2.

Proof of Theorem 2 We expand the Rademacher complexity in (5) using latent tensors  $\mathcal{W}^{(1)}, \ldots, \mathcal{W}^{(G)}$  for the reshaped latent tensor nuclear norm as

$$R_{\mathbf{S}}(l\circ(\mathcal{W}^{(1)}+\cdots+\mathcal{W}^{(G)})) = \frac{1}{|\mathbf{S}|} \mathbb{E}_{\sigma} \left[ \sup_{\mathcal{W}^{(1)}+\cdots+\mathcal{W}^{(G)}=\mathbf{W}\in\mathbf{W}_{r_{1}}} \sum_{i_{1},\ldots,i_{K}} \Sigma_{i_{1},\ldots,i_{K}} l(\mathcal{X}_{i_{1},\ldots,i_{K}},\mathcal{W}_{i_{1},\ldots,i_{K}}) \right],$$

where  $\Sigma_{i_1,...,i_K} = \sigma_j$  when  $(i_1, ..., i_K) \in S$  and  $\Sigma_{i_1,...,i_K} = 0$ , otherwise. We analyze the Rademacher complexity as

$$R_{S}\left(lo\left(\mathcal{W}^{(1)} + \dots + \mathcal{W}^{(G)}\right)\right) = \frac{1}{|S|} \mathbb{E}_{\sigma} \left[\sup_{\mathcal{W}^{(1)} + \dots + \mathcal{W}^{(G)} = \mathcal{W}_{e}W_{r1}} \sum_{i_{1},\dots,iK} \Sigma_{i_{1},\dots,iK} l\left(\mathcal{X}_{i_{1},\dots,iK}, \mathcal{W}_{i_{1},\dots,iK}\right)\right],$$

$$\leq \frac{\Lambda}{|S|} \mathbb{E}_{\sigma} \left[\sup_{\mathcal{W}^{(1)} + \dots + \mathcal{W}^{(G)} = \mathcal{W}_{e}W_{r1}} \sum_{i_{1},\dots,iK} \Sigma_{i_{1},\dots,iK} \mathcal{W}_{i_{1},\dots,iK}\right], \quad (\text{Rademacher contraction})$$

$$\leq \frac{\Lambda}{|S|} \mathbb{E}_{\sigma} \sup_{\mathcal{W}^{(1)} + \dots + \mathcal{W}^{(G)} = \mathcal{W}_{e}W_{r1}} \|\mathcal{W}\|_{r\_\text{latent}} \|\Sigma\|_{r\_\text{latent}*} \quad (\text{Duality relationship}). \tag{16}$$

(a) For a tensor with multilinear rank, using Lemma 4 we obtain

$$\begin{aligned} \|\mathcal{W}\|_{r\_latent} &= \inf_{\mathcal{W}^{(1)}+\dots+\mathcal{W}^{(G)}=\mathcal{W}} \sum_{g=1}^{G} \|\mathcal{W}_{(D_{1}^{(g)},\dots,D_{m_{g}}^{(g)})}^{(k)}\|_{\star} \\ &\leq \min_{g\in G} \left(\frac{\prod_{k=1}^{K} r_{k}}{\max_{j=1,\dots,M} \prod_{i\in D_{j}^{(g)}} r_{i}}\right) \gamma_{1}(\mathcal{W}_{(D_{1}^{(g)},\dots,D_{M_{g}}^{(g)})}^{(g)}). \end{aligned}$$
(17)

Using Lemma 1 we can bound  $\mathbb{E}_{\sigma} \|\Sigma\|_{r_{\text{latent}^*}}$  as

$$\mathbb{E}_{\sigma} \|\Sigma\|_{r_{\_} \text{latent}^{*}} = \max_{g \in G} \left\| \mathcal{W}_{\left(D_{1}^{(g)}, \dots, D_{M_{g}}^{(g)}\right)}^{(g)} \right\|_{\star} \le 4 \max \log(4M_{g}) \sum_{g \in G}^{M_{g}} \sqrt{\prod_{p \in D_{j}^{(g)}} n_{p}}.$$
 (18)

By substituting (17) and (18) to (16), we obtain the following bound

$$R_{\mathbf{S}}(l \circ \mathcal{W}) \leq \frac{cA}{|\mathbf{S}|} \min_{g \in G} \left( \frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\dots,M} \prod_{i \in D_j^{(g)}} r_i} \right) \gamma_1(\mathcal{W}_{(D_1^{(g)},\dots,D_{M_g}^{(g)})}^{(g)})$$
$$\max_{g \in G} \log(4M_g) \sum_{j=1}^{M_g} \sqrt{\prod_{p \in D_j^{(g)}} n_p}.$$

(b) For tensor with CP rank, using Lemma 4 we obtain

$$\|\mathcal{W}\|_{r\_latent} = \inf_{\mathcal{W}^{(1)} + \dots + \mathcal{W}^{(G)} = \mathcal{W}} \sum_{g=1}^{G} \|\mathcal{W}_{(D_{1}^{(g)}, \dots, D_{m_{g}}^{(g)})}^{(k)}\|_{\star} \le \min_{g \in G} r_{cp} \gamma_{1} \left(\mathcal{W}_{(D_{1}^{(g)}, \dots, D_{M_{g}}^{(g)})}^{(g)}\right).$$
(19)

By substituting (19) and (18) to (16), we obtain the following bound

$$R_{S}(l \circ \mathcal{W}) \leq \frac{c\Lambda}{|S|} \min_{g \in G} r_{cp} \gamma_{1}(\mathcal{W}_{(D_{1}^{(g)}, \dots, D_{M_{g}}^{(g)})}^{(g)}) \max_{g \in G} \log(4M_{g}) \sum_{j=1}^{M_{g}} \sqrt{\prod_{p \in D_{j}^{(g)}} n_{p}}.$$

Finally, we derive the Rademacher complexity for the tensor completion model regularized by the Schatten TT norm.

**Theorem 3** Consider a K-mode tensor  $W \in \mathbb{R}^{n_1 \times \ldots \times n_K}$  with a multilinear rank of  $(r_1, \ldots, r_K)$ . Let us consider the hypothesis class  $W_{TT} = \{W | ||T||_{s,T} \le t\}$ . Then Rademacher complexity is bounded as

$$R_{\rm S}(l\circ\mathcal{W}) \le \frac{c'A}{|{\rm S}|} \sum_{k=1}^{K-1} \min\bigg(\prod_{i=1}^k \sqrt{r_i}, \prod_{j=k+1}^K \sqrt{r_j}\bigg) B_{\mathcal{T}_{k=1,\ldots,K-1}}\bigg(\sqrt{\prod_{i< k} n_i} + \sqrt{\prod_{j\ge k}^K n_j}\bigg),\tag{20}$$

where  $\|\mathcal{W}\|_{\mathrm{F}} \leq B_{\mathcal{T}}$  and c' is a constant.

**Proof** For this case we consider the hypothesis class  $W_{TT}$  for the Rademacher complexity follows as

$$R_{\rm S}(l\circ\mathcal{W}) = \frac{1}{|{\rm S}|} \mathbb{E}_{\sigma} \left[ \sup_{\mathcal{W}_{\rm TT}\in {\rm W}} \sum_{i_1,\ldots,i_K} \Sigma_{i_1,\ldots,i_K} l(\mathcal{X}_{i_1,\ldots,i_K},\mathcal{W}_{i_1,\ldots,i_K}) \right],$$

where  $\Sigma_{i_1,...,i_K} = \sigma_j$  when  $(i_1, ..., i_K) \in S$  and  $\Sigma_{i_1,...,i_K} = 0$ , otherwise. Now we analyze the Rademacher complexity for the hypothesis class  $W_{\text{TT}}$ . We have

$$R_{S}(l \circ \mathcal{W}) = \frac{1}{|S|} \mathbb{E}_{\sigma} \left[ \sup_{\mathcal{W} \in W_{TT}} \sum_{i_{1},...,i_{K}} \Sigma_{i_{1},...,i_{K}} l(\mathcal{X}_{i_{1},...,i_{K}}, \mathcal{W}_{i_{1},...,i_{K}}) \right],$$

$$\leq \frac{\Lambda}{|S|} \mathbb{E}_{\sigma} \left[ \sup_{\mathcal{W} \in W_{TT}} \sum_{i_{1},...,i_{K}} \Sigma_{i_{1},...,i_{K}} \mathcal{W}_{i_{1},...,i_{K}} \right], \quad (\text{Rademacher contraction}) \quad (21)$$

$$\leq \frac{\Lambda}{|S|} \mathbb{E}_{\sigma} \sup_{\mathcal{W} \in W_{TT}} ||W||_{s,T} ||\Sigma||_{s,T^{*}}, \quad (\text{Duality relationship})$$

where  $\|\cdot\|_{s,T^*}$  is the dual norm of  $\|\cdot\|_{s,T}$ . The last step can be obtained by applying the Holder's inequality to the sum of trace norms in the Schatten TT norm.

Considering  $\|\mathcal{W}\|_{s,T}$ , we can expand it as

$$\|\mathcal{W}\|_{s,T} = \frac{1}{K-1} \sum_{k=1}^{K-1} \|Q_k(\mathcal{I})\|_{\mathrm{tr}} = \frac{1}{K-1} \sum_{k=1}^{K-1} \sum_{i_k=1}^{\hat{r}_k} \gamma_{i_k}(Q_k(\mathcal{I})),$$

where  $Q_k : \mathcal{T} \to \mathbb{R}^{n_{\geq k} \times n_{k<}}$  is a reshaping operator, and  $\gamma_{i_k}()$  and  $\hat{r}_k$  are the  $i_k$ th singular value and the rank of the reshaped tensor by  $Q_k$ , respectively. Using the Cauchy-Schwarz inequality, we have

$$\|\mathcal{W}\|_{s,T} \leq \frac{1}{K-1} \sum_{k=1}^{K-1} \sqrt{\hat{r}_k} \sqrt{\sum_{i_k=1}^{\hat{r}_k} \gamma_{i_k}^2} (Q_k(\mathcal{T})) = \frac{1}{K-1} \sum_{k=1}^{K-1} \sqrt{\hat{r}_k} B_{\mathcal{T}},$$

where  $||\mathcal{T}||_{\rm F} = B_{\mathcal{T}}$ . Using Lemmas 1 and 2, we can infer that

$$\|\mathcal{W}\|_{s,T} \le \frac{1}{K-1} \sum_{k=1}^{K-1} \min\left(\prod_{i=1}^{k} \sqrt{r_i}, \prod_{j=k+1}^{K} \sqrt{r_j}\right) B_{\mathcal{T}},$$
(22)

Similar to the overlapped trace norm (Tomioka and Suzuki 2013), the Schatten TT norm also sums nuclear norms of the the same tensor reshaped into different matrices. Hence, we can extend the dual norm of the overlapped trace norm in Tomioka and Suzuki (2013) to the Schatten TT norm. Using (Tomioka and Suzuki 2013), it is easy to the dual norm of Schatten TT norm as

$$\|\Sigma\|_{s,T^*} = \inf_{\Sigma^{(1)} + \dots + \Sigma^{(k)} = \Sigma} \sum_{k=1}^{k-1} \|Q_k(\Sigma^{(k)})\|_{\text{op}}$$

We want to bound

$$\mathbb{E} \|\Sigma\|_{s,T^*} = \mathbb{E} \inf_{\Sigma^{(1)} + \dots + \Sigma^{(K)} = \Sigma} \sum_{k=1}^{K-1} \left\| Q_k(\Sigma^{(k)}) \right\|_{\text{op}}$$

and since we can take any of  $\Sigma^{(k)}$ , k = 1, ..., K to be equal to  $\Sigma$ , we have

$$\mathbb{E} \|\Sigma\|_{s,T^*} \leq \min_{k=1,\ldots,K-1} \|Q_k(\Sigma)\|_{\text{op}}.$$

We apply Latała's Theorem (Latała 2005; Shamir and Shalev-Shwartz 2014) to the reshaping by the  $Q_k$  operator and bound  $\mathbb{E} \|Q_k(\Sigma)\|_{op}$  as

$$\mathbb{E} \left\| \mathcal{Q}_{k}(\Sigma) \right\|_{\text{op}} \leq C_{1} \left( \sqrt{\prod_{i < k} n_{i}} + \sqrt{\prod_{j \ge k}^{K} n_{j}} + \sqrt[4]{|\mathcal{Q}_{k}(\Sigma)|} \right),$$
  
and since  $\sqrt[4]{|\mathcal{Q}_{k}(\Sigma)|} \leq \sqrt[4]{\prod_{i=1}^{K} n_{i}} \leq \frac{1}{2} \left( \sqrt{\prod_{i < k} n_{i}} + \sqrt{\prod_{j \ge k}^{K} n_{j}} \right),$  we have,  
$$\mathbb{E} \left\| \mathcal{Q}_{k}(\Sigma) \right\|_{\text{op}} \leq \frac{3C_{1}}{2} \left( \sqrt{\prod_{i < k} n_{i}} + \sqrt{\prod_{j \ge k}^{K} n_{j}} \right).$$

This gives us the bounds for  $\mathbb{E} \|\Sigma\|_{s,T^*}$  as

$$\mathbb{E} \|\Sigma\|_{s,T^*} \le \min_{k=1,\dots,K-1} \frac{3C_1}{2} \left( \sqrt{\prod_{i< k} n_i + \sqrt{\prod_{j\ge k} n_j}} \right).$$
(23)

By combining (22) and (23) to (21), we obtain

$$R_{\rm S}(l\circ\mathcal{W}) \leq \frac{c'\Lambda}{|{\rm S}|(K-1)} \sum_{k=1}^{K-1} \min\left(\prod_{i=1}^{k} \sqrt{r_i}, \prod_{j=k+1}^{K} \sqrt{r_j}\right) B_{\mathcal{T}} \min_k\left(\sqrt{\prod_{i$$

## References

- Berclaz, J., Fleuret, F., Turetken, E., & Fua, P. (2011). Multiple object tracking using K-shortest paths optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33, 1806–19.
- Carroll, J. D., & Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrika*, 35(3), 283–319.
- El-Yaniv, R., & Pechyony, D. (2007). Transductive rademacher complexity and its applications. *Learning Theory*, 4539, 157–171.
- Fazel, M., Hindi, H., & Boyd, S. P. (2001). A rank minimization heuristic with application to minimum order system approximation. In *Proceedings of the 2001 American Control Conference*. (Cat. No.01CH37148) (Vol. 6, pp. 4734–4739).
- Guo, X., Yao, Q., & Kwok, J. T. (2017). Efficient sparse low-rank tensor completion using the frank-wolfe algorithm. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4–9, 2017, San Francisco, California, USA. (pp. 1948–1954).
- Hackbusch, W. (2012). Tensor spaces and numerical tensor calculus. Berlin Heidelberg: Springer Series in Computational Mathematics. Springer. ISBN 9783642280276.
- Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an explanatory multimodal factor analysis. UCLA Working Papers in Phonetics, 16, 1–84.
- Hillar, C. J., & Lim, L.-H. (2013). Most tensor problems are np-hard. Journal of ACM, 60(6), ISSN 0004-5411.
- Hitchcock, F. L. (1927). The expression of a tensor or a polyadic as a sum of products. Journal of Mathematics and Physics, 6(1), 164–189.
- Imaizumi, M., Maehara, T., & Hayashi, K. (2017). On tensor train rank minimization: Statistical efficiency and scalable algorithm. In NIPS, pp. 3933–3942.

- Karatzoglou, A., Amatriain, X., Baltrunas, L., & Oliver, N. (2010). Multiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering. In *RecSys* (pp. 79–86). ACM.
- Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. SIAM Review, 51(3), 455–500.
- Latała, R. (2005). Some estimates of norms of random matrices. Proceedings of the American Mathematical Society, 133(5), 1273–1282.
- Lim, L., & Comon, P. (2014). Blind multilinear identification. *IEEE Transactions on Information Theory*, 60(2), 1260–1280.
- Liu, J., Musialski, P., Wonka, P., & Ye, J. (2009). Tensor completion for estimating missing values in visual data. In *ICCV* (pp. 2114–2121).
- Mu, C., Huang, B., Wright, J., & Goldfarb, D. (2014). Square deal: Lower bounds and improved relaxations for tensor recovery. In *ICML* (pp. 73–81).
- Oseledets, I. V. (2011). Tensor-train decomposition. SIAM Journal on Scientific Computing, 33(5), 2295– 2317, ISSN 1064-8275.
- Rai, P., Hu, C., Harding, M., & Carin, L. (2015). Scalable probabilistic tensor factorization for binary and count data. IJCAI'15, pp. 3770–3776. AAAI Press. ISBN 978-1-57735-738-4.
- Raskutti, G., Chen, H., Yuan, M. (2015). Convex regularization for high-dimensional multi-response tensor regression. CoRR, abs/1512.01215v2.
- Shamir, O., & Shalev-Shwartz, S. (2014). Matrix completion with the trace norm: Learning, bounding, and transducing. *Journal of Machine Learning Research*, 15, 3401–3423.
- Song, Q., Ge, H., Caverlee, J., & Hu, X. (2017). Tensor completion algorithms in big data analytics. CoRR, abs/1711.10105.
- Tomioka, R., & Suzuki, T. (2013). Convex tensor decomposition via structured schatten norm regularization. In NIPS.
- Wimalawarne, K., Sugiyama, M., & Tomioka, R. (2014). Multitask learning meets tensor factorization: Task imputation via convex optimization. In *NIPS*.
- Yang, Y., Feng, Y., & Suykens, J. A. K. (2015). A rank-one tensor updating algorithm for tensor completion. IEEE Signal Processing Letters, 22(10), 1633–1637. ISSN 1070-9908.
- Yuan, M., & Zhang, C.-H. (2016). On tensor completion via nuclear norm minimization. Foundations of Computational Mathematics, 16(4), 1031–1068.
- Zheng, V. W., Cao, B., Zheng, Y., Xie, X., & Yang, Q. (2010). Collaborative filtering meets mobile recommendation: A user-centered approach. In AAAI, AAAI'10, pp. 236–241. AAAI Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.